DOI: http://dx.doi.org/10.5902/2179460X11216
Revista do Centro do Ciências Naturais e Exatas - UFSM
Ciência e Natura, Santa Maria,
ISSN: 2179-460X, v. 35 n. 1 jul. 2013, p. 098-105

Comparação de técnicas estatísticas para analisar a relação entre doenças respiratórias e concentrações de poluentes atmosféricos.

Comparison of statistical techniques to analyze the relation between breathing diseases and concentrations of atmospheric pollutants.

Angela Radünz Lazzari

Departamento de Matemática - UNISC

Resumo

A poluição atmosférica é um fator de risco para a saúde da população. Seus efeitos nocivos na população são observados mesmo quando os poluentes atmosféricos encontram-se dentro dos parâmetros previstos em legislação específica e se desenvolvem, principalmente, por meio de doenças respiratórias. O objetivo deste trabalho foi analisar as relações existentes entre as concentrações de poluentes atmosféricos e incidências de doenças respiratórias na cidade de Porto Alegre, em 2005 e 2006. Foram utilizadas as análises de regressão linear múltipla, regressão logística ordinal e modelos lineares generalizados. Os resultados apresentados mostram bom ajustamento pelas três técnicas. A regressão logística ordinal detectou, apenas, influência positiva da temperatura do ar e da umidade relativa do ar nas internações por doenças respiratórias. A regressão linear múltipla relacionou as internações com as variáveis meteorológicas, negativamente, e com material particulado (PM_{10}), positivamente. O modelo linear generalizado detectou influência negativa das variáveis meteorológicas e positiva dos poluentes, ozônio troposférico (O_3) e PM_{10}, nas internações. Comparando-se as três técnicas estatísticas para analisar o mesmo conjunto de dados, pode-se concluir que todas elas apresentaram um modelo com bom ajuste aos dados, mas a técnica de modelos lineares generalizados apresentou maior sensibilidade em captar a influência dos poluentes, o que não ocorreu na regressão logística ordinal e na regressão linear múltipla.

Palavras-chave: doenças respiratórias, poluição atmosférica, regressão linear, regressão logística ordinal, modelo linear generalizado.

Abstract

Air pollution is a risk factor for the population health. Its harmful effects on the population are observed even when the atmospheric pollutants are within the parameters set out in specific legislation, and they develop mainly through respiratory diseases. The aim of this study was to analyze the relationship between the concentrations of air pollutants and the incidence of respiratory diseases in the city of Porto Alegre, in 2005 and 2006. Applied multiple linear regression analysis, ordinal logistic regression and generalized linear models were used in the work. The results show good adjustment by the three techniques. The ordinal logistic regression detected only positive influence of air temperature and relative humidity in hospitalizations for respiratory diseases. Multiple linear regression related negatively hospitalizations with meteorological variables and positively with the particulate matter (PM_{10}). The generalized linear model detected negative influence of meteorological variables and positive of pollutants, tropospheric ozone (O_3) and PM_{10} in hospitalizations. Comparing the three statistical techniques to analyze the same data set, it can be concluded that all of them had a model with good fit to the data, but the technique of generalized linear models showed higher sensitivity in capturing the influence of pollutants, except in ordinal logistic regression and multiple linear regression.

Keywords: Respiratory diseases, air pollution, linear regression, ordinal logistic regression, generalized linear model

Recebido em: 15.02.12, Revisado em: 10.06.13, Aceito em: 01.07.13

1. INTRODUÇÃO

A poluição atmosférica, em grandes centros urbanos, representa um fator de risco para a saúde da população. As atividades antrópicas são responsáveis por alterações e degradações ambientais. As indústrias e os veículos automotores possuem grande potencial poluidor da atmosfera. Nas cidades industrializadas, a emissão de gases tóxicos e de partículas na atmosfera a poluição provocada pela circulação de veículos, gerando muitas vezes situações críticas para a saúde da população. Os efeitos nocivos da poluição do ar são observados mesmo quando os poluentes atmosféricos encontram-se dentro dos parâmetros previstos em legislação específica.

Dos poluentes presentes na atmosfera, pode-se destacar o ozônio troposférico (O_3) e o material particulado (PM_{10}), pois são poluentes monitorados por órgãos ambientais e podem causar problemas à saúde humana.

O ozônio troposférico (O_3), gás composto por três átomos de oxigênio, não é emitido diretamente, mas se forma na baixa atmosfera através de reações fotoquímicas (catalisadas pelos raios ultravioletas do sol) entre os óxidos de nitrogênio (NO_x) e os compostos orgânicos voláteis (COVs). Os picos de ozônio ocorrem tipicamente em períodos de calor, tempo seco e algumas vezes em condições atmosféricas estagnantes. O ozônio é um oxidante muito forte, citotóxico (tóxico às células) e que, mesmo em baixas concentrações, pode causar agravamento de doenças respiratórias. Pela Resolução n° 03 de 28/06/1990 do Conselho Nacional do Meio Ambiente (CONAMA), o padrão de qualidade do ar para as concentrações de O_3 deve ser abaixo de 160 $\mu g\ m^{-3}$. O Material Particulado (PM) é o termo usado para partículas suspensas no ar na forma de poeiras, fumos, fumaça ou névoas (partículas líquidas). Sua composição e tamanho dependem da fonte de emissão, mas, em geral, o PM é constituído por componentes múltiplos, tais como: sulfatos, sais, metais, partículas de carbono e sílica finamente divididas. As classificações usuais são partículas totais em suspensão e partículas inaláveis. As primeiras são partículas cujo tamanho é, em geral, inferior a 100 micra. O segundo grupo compreende as partículas inferiores ou iguais a 10 µm (PM_{10}) e partículas inferiores a 2,5 µm ($PM_{2,5}$). As partículas inaláveis podem atingir as vias respiratórias transportando os gases adsorvidos em sua superfície até as porções mais distais das vias aéreas prejudicando as trocas gasosas no pulmão e podendo causar problemas respiratórios, pulmonares, asma e alergia. Pela Resolução n° 03 de 28/06/1990 do CONAMA, o padrão de qualidade do ar para as concentrações de PM_{10} deve ser inferior a 150 $\mu g\ m^{-3}$ (FEPAM, 2002; Branis et al, 2009; York et al, 2009; Bailey e Solomon, 2004).

Na literatura, têm-se alguns trabalhos que relacionam dados de morbidade por doenças respiratórias com os efeitos da poluição atmosférica. Os autores DOMINICI et al (2006), MARTINS et al (2002) e SCHWARTZ (1996) investigaram esses efeitos em idosos. DOMINICI et al (2006) analisaram as internações hospitalares por doenças respiratórias e cardiovasculares, em pessoas com mais de 65 anos, associadas à exposição ao $PM_{2,5}$, em 204 municípios com mais de 200.000 habitantes, em várias regiões dos Estudos Unidos. Utilizando modelos hierárquicos e de regressão de Poisson, concluíram que a exposição, à curto prazo, às partículas finas aumenta o risco de internações por doenças respiratórias e cardiovasculares. MARTINS et al (2002) analisaram a relação entre poluentes e morbidade por pneumonia e por gripe entre 1996 e 1998. Utilizaram modelo aditivo generalizado de regressão de Poisson e concluíram que os poluentes O_3 e SO_2 estão diretamente associados à pneumonia e à gripe. SCHWARTZ (1996) analisou a relação entre concentrações de poluentes atmosféricos (PM_{10} e O_3) com internações hospitalares por doenças respiratórias, em pessoas com idade acima de 65 anos, em Spokane, WA, Estados Unidos. Utilizou a análise de regressão de Poisson e concluiu que os poluentes PM_{10} e O_3 estão associados positivamente com as internações por doenças respiratórias.

Outros estudos mostram resultados com crianças e idosos. GOUVEIA et al (2003) estudaram os efeitos nocivos da poluição do ar na saúde da população nas cidades de São Paulo e Rio de Janeiro. Com a utilização de modelos lineares, concluiu-se que o aumento nos níveis de poluentes atmosféricos esteve associado ao aumento na mortalidade e internações hospitalares por causas respiratórias e cardiovasculares em crianças e idosos em ambos os municípios.

Outros estudos especificamente com crianças. SEGALA et al (2008) realizaram estudo na região de Paris, de 1997 a 2001, em períodos de inverno em que verificaram a existência de associação entre ocorrência de bronquiolite, em crianças menores de 3 anos, e poluição do ar. Foi utilizado o modelo aditivo generalizado, resultando em associações positivas entre os poluentes, consultas e internações hospitalares pela doença. BAKONYI et al (2004) investigaram os efeitos causados pela poluição atmosférica na morbidade por doenças respiratórias em crianças entre 1999 e 2000 na cidade de Curitiba, PR. Foi utilizado o modelo aditivo generalizado de regressão de Poisson e foi verificado verificando

que todos os poluentes apresentaram efeitos sobre as doenças respiratórias de crianças.

CONCEIÇÃO et al (2001) fazem uma descrição e comparação de duas classes de modelos - os Modelos Lineares Generalizados (MLG) e os Modelos Aditivos Generalizados (MAG) - que podem ser utilizadas para avaliar a associação entre poluição atmosférica e marcadores de morbi-mortalidade. Como exemplo de aplicação, foi avaliada a associação entre mortalidade em idosos e poluição atmosférica na cidade de São Paulo no período de 1994 a 1997. As duas classes de modelos produziram resultados coerentes, mas os modelos estatisticamente mais sofisticados tiveram mais poder para detectar efeitos significantes. Foram observadas associações entre mortalidade e os níveis de CO, SO_2 e, em menor escala, PM_{10}. As associações observadas foram dose-dependente e evidentes após um curto período de exposição.

O objetivo deste trabalho está em analisar as relações existentes entre as incidências de doenças respiratórias e as concentrações de poluentes atmosféricos na cidade de Porto Alegre com dados de 2005 e 2006. Utilizando técnicas estatísticas de dependência para analisar esta relação, onde a incidência de doenças respiratórias, medida pelo número de internações mensais pelo SUS, é considerada a variável dependente e as demais variáveis independentes.

2. MATERIAL E MÉTODOS

Os dados utilizados nas análises se referem ao município de Porto Alegre para os anos de 2005 e 2006. A variável que identifica a morbidade corresponde ao número de internações mensais (INTERN) por doenças respiratórias pelo Sistema Único de Saúde (SUS). Os dados meteorológicos foram fornecidos pelo Instituto Nacional de Meteorologia (Inmet) que correspondem às temperaturas médias mensais do ar (TEMP) e à umidade relativa média mensal do ar (UMID). Os dados referentes aos poluentes atmosféricos foram fornecidos pela Fundação Estadual de Proteção Ambiental (Fepam) que são as concentrações médias e máximas mensais de PM_{10} e O_3.

Para analisar os efeitos nocivos da poluição atmosférica na saúde da população, buscaram-se trabalhos já realizados nessa área. Muitos deles relacionaram a poluição atmosférica com doenças respiratórias, principalmente nas faixas etárias de crianças e idosos. A maioria dos trabalhos utilizou modelos aditivos generalizados e regressão de Poisson. No presente trabalho são apresentadas as seguintes técnicas estatísticas, para analisar os efeitos na saúde da população por poluentes atmosféricos, que são: a análise de regressão linear múltipla, análise de regressão logística ordinal e modelos lineares generalizados.

As análises estatísticas envolvem, primeiramente, análises descritivas das variáveis e a aplicação de técnicas de análise multivariável. A análise multivariável se refere a todas as técnicas estatísticas que simultaneamente analisam múltiplas medidas sobre indivíduos ou objetos sob investigação, ou seja, qualquer análise simultânea de mais do que duas variáveis pode ser considerada, a princípio, como multivariável. As técnicas a serem utilizadas neste trabalho são todas técnicas de dependência, onde são classificadas por terem uma ou mais variáveis dependentes e diversas variáveis independentes.

A análise de regressão múltipla envolve uma variável dependente quantitativa (Y) relacionada com k variáveis independentes quantitativas (X_i), com o objetivo de encontrar um modelo matemático que relacione e explique o comportamento da variável dependente. A função de regressão múltipla pode ser descrito da seguinte forma:

$$Y = \beta_0 + \beta_1.X_1 + ... + \beta_k.X_k$$

Em que:

Y é a variável dependente;

X_1, ... X_k são as variáveis independentes;

β_0, ... β_k são denominados parâmetros da regressão.

Na análise de regressão, a multicolinearidade tem substancial impacto sobre a especificação final do modelo. Como por exemplo, quando duas variáveis independentes são altamente correlacionadas e tem correlações quase iguais com a variável dependente. O critério para inclusão ou eliminação nessas abordagens é maximizar o poder preditivo incremental da variável adicional. Se uma dessas variáveis entrar no modelo será muito improvável que a outra também entre, pois existe pouca variância individual para cada variável separadamente. Portanto, deve-se avaliar os efeitos da multicolinearidade na interpretação do modelo examinando não apenas a equação de regressão final, mas também as correlações diretas de todas as variáveis independentes potenciais (Hair et al, 2009; Corrar et al, 2009).

Para avaliar se o modelo se ajusta adequadamente aos dados, é feita uma análise de diagnóstico,

examinando os resíduos. Eles devem seguir uma distribuição normal $N(0,\sigma^2)$. No gráfico dos resíduos versus variável dependente, os resíduos devem se distribuir aleatoriamente em torno da média zero, sem nenhum padrão específico dos resíduos (Hair et al, 2009; Agresti e Finlay, 2012).

Os modelos de regressão logística representam uma combinação de regressão múltipla e análise discriminante múltipla em que se utiliza uma ou mais variáveis independentes para prever uma única variável dependente categórica.

Conforme ABREU et al (2009) no modelo de regressão logística ordinal é definido pela equação:

$$g(p_j) = \alpha_j + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k,$$

Em que:

$g(p_j)$ é a função de ligação: $g(p_j)=\ln(p_j/1-p_j)$;

$X_1, \dots X_k$ são as variáveis independentes;

α_j refere-se ao intercepto do modelo;

$\beta_1, \dots \beta_k$ são parâmetros que descrevem os efeitos das variáveis independentes na variável dependente (Y);

Y é a variável dependente com n-categorias codificadas em 1, 2, ..., n. As n categorias de Y condicionalmente aos valores de X_i ocorrem com probabilidades $p_1, p_2, \dots p_n$, isto é, $p_j = P(Y=j)$, para j = 1, 2, ... n.

A variável dependente "número de internações" foi agrupada, dando origem a uma variável ordinal. O agrupamento foi feito utilizando os percentis 33 e 67, formando assim três grupos (n=3): o primeiro indicando um número baixo, o segundo indicando um número médio e o terceiro indicando um número alto de internações.

Existe na literatura uma série de modelos de regressão logística ordinal, sendo que o modelo aplicado neste trabalho foi o Modelo de Chances Proporcionais (MCP). Este, também chamado de modelo do logit (função distribuição logística padrão cumulativa), produz estimativas de simples compreensão. Seu uso é indicado quando a variável dependente era originalmente uma variável contínua e que posteriormente foi agrupada. Esse modelo compara a probabilidade de uma resposta igual ou menor a uma determinada categoria (*j = 1, 2, ..., n-1*), com probabilidade de uma resposta maior que esta categoria. Além disso, o modelo é composto por *n-1* equações lineares paralelas. No caso particular de apenas duas categorias (*n*=2), o MCP corresponde exatamente ao tradicional modelo de regressão logística binária.

O termo α_j varia para cada uma das n categorias e cada β não depende do índice j, implicando que a relação entre X_i e Y é independente da categoria. Esse modelo possui uma suposição de Odds proporcionais acerca dos (n-1) pontos de corte, também chamada de suposição de regressão paralela, que é assumida para cada variável independente incluída no modelo. Essa suposição deve ser testada para cada variável independente separadamente e no modelo final (Agresti e Finlay, 2012; Abreu et al, 2009).

Os modelos lineares generalizados (MLG), de acordo com MCCULLAGH e NELDER (1989) e AGRESTI e FINLAY (2012), constituem uma classe ampla de modelos que inclui modelos de regressão comum para variáveis dependentes com distribuição normal, modelos alternativos para variáveis contínuas que não presumem normalidade e modelos para variáveis dependentes discretas ou categóricas. Em variáveis de frequências, como é o caso do "número de internações", a distribuição mais aproximada é a distribuição de Poisson.

Em um MLG, represente $\mu=E(y)$. μ varia de acordo com os valores das variáveis explicativas, que entram linearmente como previsoras no lado direito da equação do modelo. A fórmula do MLG estabelece que: $g(\mu)=\alpha+\beta_1x_1+\beta_2x_2+\dots+\beta_kx_k$.

A função $g(\mu)$ é chamada de função de ligação porque liga a média da variável resposta às variáveis explicativas. Foi considerada a função de ligação $g(\mu)=\log(\mu)$ por ser a função de ligação canônica do MLG com a distribuição de Poisson.

As análises estatísticas foram realizadas com auxílio da planilha eletrônica Excel e dos softwares estatísticos Minitab e Statistica. O nível de significância utilizado foi de 0,05.

3. RESULTADOS E DISCUSSÃO

A maioria dos trabalhos que analisam a relação entre as concentrações de poluentes e suas consequências com a saúde, geralmente, doenças respiratórias, analisa as faixas etárias de crianças e de idosos por serem idades consideradas mais sensíveis a essas doenças. Neste trabalho, optou-se por analisar as internações em geral, mas também as internações de crianças e de idosos em separado para identificar diferenças. Para os dois anos em estudo, 2005 e 2006, verificou-se uma média de 24% das internações por doenças respiratórias em crianças (de 0 a 4 anos) e 23% das internações são compostas de pessoas

com 65 anos ou mais. Nas análises que se seguem, observou-se que as internações contendo todas as idades apresentou melhor ajustamento que as outras faixas etárias, portanto optou-se por analisar apenas o grupo geral de internações.

Tabela 1 – Correlação linear de Pearson.

	INTERN	*INTERN CRIANÇAS*	*INTERN IDOSOS*	*TEMP*	*UMID*	*PM_{10}*
INTERN CRIANÇAS	*0,972 (<0,001)*					
INTERN IDOSOS	*0,886 (<0,001)*	*0,820 (<0,001)*				
TEMP	*-0,803 (<0,001)*	*-0,683 (<0,001)*	*-0,683 (<0,001)*			
UMID	*0,229 (0,282)*	*0,148 (0,491)*	*0,148 (0,491)*	*-0,594 (0,002)*		
PM_{10}	*0,144 (0,501)*	*0,136 (0,526)*	*0,136 (0,526)*	*-0,023 (0,917)*	*0,177 (0,409)*	
O_3	*-0,435 (0,034)*	*-0,440 (0,032)*	*-0,376 (0,070)*	*0,624 (0,001)*	*-0,395 (0,056)*	*-0,113 (0,600)*

Nota: Conteúdo das células: Correlação de Pearson (p-valor)

Na Tabela 1, são apresentadas as correlações entre as variáveis em estudo, com o respectivo p–valores, do teste H0: correlação populacional = 0, considerando correlação significativa aquela com p≤0,05. Pode-se observar que a temperatura média do ar se correlaciona negativamente com todas as variáveis de internações. Já a umidade média do ar não se correlaciona com nenhuma das variáveis de interesse, mas apenas negativamente com a temperatura do ar. Como essa correlação não é muito alta, esta variável será mantida nas análises de regressão. O Poluente PM_{10} não apresenta correlação significativa com nenhuma das variáveis. O O_3 apresenta correlação negativa com as internações em geral e de crianças e correlação positiva com a temperatura, pois é um poluente que apresenta maiores concentrações em períodos de muito calor, e as internações são maiores no inverno.

Inicialmente utilizou-se a regressão linear múltipla com as seguintes variáveis: INTERN, como variável dependente, e TEMP, UMID, PM_{10} e O_3, como variáveis independentes. Chegou-se a seguinte equação:

INTERN = 4251– 69,3.TEMP – 25,7UMID + 1,94.PM10.

Os resultados do modelo acima estão apresentados na Tabela 2.

Tabela 2 – Modelo ajustado pela Análise de Regressão Linear Múltipla

	Coeficiente	*Erro padrão*	*t*	*P-valor*	*FIV*
Constante	*4250,6*	*716,3*	*5,93*	*<0,001*	
TEMP	*-69,341*	*8,683*	*-7,99*	*<0,001*	*1,563*
UMID	*-25,688*	*7,958*	*-3,23*	*0,004*	*1,612*
PM_{10}	*1,942*	*1,065*	*1,82*	*0,083*	*1,044*

R^2=77,6%; 74,3%.

Pela equação ajustada pode-se concluir que há uma relação inversa entre as internações, a temperatura e a umidade relativa do ar, indicando que quanto menor for a temperatura e a umidade do ar, conjuntamente com maior índice de PM_{10}, maiores são as internações.

O fator de inflação da variância (FIV) representa o aumento multiplicativo na variância do estimador devido a X_i estar correlacionada com as outras variáveis independentes. Ele indica a ausência (valores abaixo de 1) ou presença (valor acima de 10) de multicolinearidade.

Tabela 3 – Resultados das estimativas do modelo melhor ajustado aos dados pela regressão logística ordinal

Variável	*Coeficiente*	*Erro padrão*	*p-valor*	*Razão de chances*	*Intervalo de confiança a 95%*	
					LI	*LS*
Constante 1	*-54,61*	*19,61*	*0,005*			
Constante 2	*-50,83*	*18,93*	*0,007*			
TEMP	*1,17*	*0,33*	*<0,001*	*3,23*	*1,68*	*6,20*
UMID	*0,39*	*0,18*	*0,030*	*1,48*	*1,04*	*2,11*

Log-Verossimilhança = - 11,999
Teste para todos os coeficientes iguais a zero: G = 28,736, GL = 2, p-valor = <0,001

As variáveis independentes apresentadas no modelo de regressão apresentam valores baixos indicando uma colinearidade aceitável (Agresti e Finlay, 2012). Para o modelo ajustado foi feita análise dos resíduos onde verificou-se que os desvios se distribuem normalmente, com média zero, e o gráfico dos resíduos versus Y se distribuem aleatoriamente em torno da média zero e não apresentam nenhum padrão específico, concluindo assim que a equação apresenta bom ajuste as dados.

Os resultados do melhor ajustamento para a análise de regressão logística ordinal são mostrados na tabela 3.

Os coeficientes das variáveis independentes apresentam estimativas significativas. A análise de regressão logística ordinal é composta por múltiplas equações de regressão logística binária, uma para cada valor de resposta, menos um. As equações são usadas para examinar como a probabilidade de algum resultado depende de mudanças nas variáveis independentes. Cada equação representa uma regressão logística binária, com dois resultados: baixo="valor da categoria ou abaixo" e alto="valor acima da categoria". A variável dependente INTERN foi agrupada, dando origem a uma variável ordinal com 3 grupos: o primeiro indicando um número baixo, o segundo indicando um número médio e o terceiro indicando um número alto de internações. Portanto, aqui resultou em duas equações:

$$g(\hat{p}_1) = -54{,}61 + 1{,}17.\ TEMP + 0{,}39.\ UMID,$$

$$g(\hat{p}_2) = -50{,}83 + 1{,}17.\ TEMP + 0{,}39.\ UMID.$$

A primeira equação produz a probabilidade de ocorrer baixa internação em um mês. A segunda equação produz a probabilidade de ocorrer baixa ou média internação em um mês. Pelos coeficientes apresentados nos modelos, pode-se dizer que há evidências de que a temperatura média do ar e a umidade relativa média do ar exercem influência positiva nas baixas internações. Maiores temperaturas e umidade relativa do ar representam um número maior de baixas internações.

A razão de chances (odds ratio) representa o incremento que a variável independente apresenta na variável dependente. A cada grau de temperatura que aumenta, eleva em 3,23 vezes a chance de ocorrer baixo número de internação. A cada ponto percentual na umidade relativa do ar que aumenta, eleva em 1,48 vezes a chance de ocorrer baixo número de internação. Por este modelo, apenas a temperatura e a umidade do ar mostraram influências significativas nas internações, confirmando o resultado da análise de regressão, com exceção do poluente PM_{10} que aqui não apresentou influência significativa.

Os indicadores da qualidade do ajuste (Goodness-of-Fit Tests) são apresentados por dois métodos, Pearson e Deviance, onde se verifica que ambos apresentam ótimos ajustes com probabilidades 0,968 e 0,994, respectivamente.

As medidas de associação, apresentadas na Tabela 4, indicam a relação entre as respostas observadas e as probabilidades previstas. A tabela mostra que 92,7% dos pares foram concordantes e apenas 7,3% discordantes, indicando uma boa capacidade preditiva do modelo.

Tabela 4 – Medidas de associação entre a variável resposta e as probabilidades preditoras.

Pares	*Número*	*%*	*Medidas resumo*	
Concordantes	*178*	*92,7*	*Somers' D*	*0,85*
Discordantes	*14*	*7,3*	*Goodman-Kruskal Gamma*	*0,85*
Relacionadas	*0*	*0,0*	*Kendall's Tau-a*	*0,59*
Total	*192*	*100,0*		

A Tabela 5 apresenta as estimativas para os parâmentros do modelo linear generalizado para a distribuição de Poisson, com função de ligação log, utilizando a variável INTERN como variável dependente e as demais como independentes. O modelo com todas as variáveis inclusas foi considerado o melhor modelo. Para esta escolha, foi feita a comparação com os demais modelos excluindo-se uma ou mais variáveis. Aqui, todas as estimativas apresentadas são significativas. Neste modelo pode-se observar a influência positiva de ambos os poluentes nas internações, confirmando resultados apresentados pelos autores citados. Quanto maiores as concentrações de poluentes e menores as temperaturas e umidade relativa do ar, maiores serão as internações por doenças respiratórias.

Tabela 5 – Estimativas para os parâmetros do Modelo Linear Generalizado

	Coeficiente	*Erro Padrão*	*Estatística Wald*	*LC Inferior 95 %*	*LC Superior 95 %*	*p*
Constante	*9,612*	*0,205*	*2195,865*	*9,210*	*10,014*	*<0,0001*
PM_{10}	*0,006*	*0,001*	*42,480*	*0,004*	*0,008*	*<0,0001*
TEMP	*-0,067*	*0,002*	*795,689*	*-0,071*	*-0,062*	*<0,0001*
UMID	*-0,021*	*0,002*	*94,128*	*-0,026*	*-0,017*	*<0,0001*
O_3	*0,006*	*0,002*	*7,613*	*0,002*	*0,010*	*0,0058*

Foi feita análise dos resíduos para o modelo ajustado, e pode-se verificar que os desvios se distribuem normalmente, com média zero. Para o diagrama de dispersão dos resíduos versus a variável dependente Y pode-se observar que os pontos se distribuem aleatoriamente em torno da média zero e não apresentam nenhum padrão específico, concluindo que a equação apresenta bom ajuste aos dados.

Comparando-se as três técnicas estatísticas para analisar o mesmo conjunto de dados, pode-se concluir que todas elas apresentaram um modelo

com bom ajuste aos dados, mas a técnica de modelos lineares generalizados apresentou maior sensibilidade em captar a influência dos poluentes, o que não ocorreu na regressão logística ordinal e na regressão linear múltipla. Analisando conjuntamente a temperatura do ar, a umidade relativa do ar e os poluentes, percebe-se uma influência negativa, tanto da temperatura do ar como da umidade relativa do ar, e uma influência positiva dos dois poluentes analisados, PM_{10} e O_3.

4. CONCLUSÃO

No presente estudo, verificou-se que as técnicas de análise apresentaram um bom ajustamento aos dados. Pela análise de regressão linear múltipla pode-se concluir que existe uma relação inversa entre as internações, as temperaturas e a umidade relativa do ar, indicando que quanto menor for a temperatura e a umidade do ar, conjuntamente com maior índice de PM_{10}, maiores são as internações. Pela regressão logística ordinal conclui-se que a temperatura média apresenta grande influência na variável dependente e a umidade média apresenta influência menor. Por este modelo, apenas a temperatura e a umidade do ar mostraram influências significativas no número de internações. Na utilização de modelos lineares generalizados teve-se um ajustamento aos dados detectando uma influência dos poluentes estudados, além das variáveis meteorológicas. Aqui, verifica-se uma influência negativa, tanto da temperatura do ar como da umidade relativa do ar, e uma influência positiva dos dois poluentes analisados, PM_{10} e O_3, nas internações por doenças respiratórias.

5. REFERÊNCIAS

ABREU, M. N. S.; SIQUEIRA, A. L.; CAIAFFA, W. T. Regressão logística ordinal em estudos epidemiológicos. **Revista de Saúde Pública**, v. 43, n. 1, p. 183-94, 2009.

AGRESTI, A.; FINLAY, B. **Métodos Estatísticos para as Ciências Sociais**. 4. Ed. – Porto Alegre: Penso, 2012.

BAILEY, D.; SOLOMON, G. Pollution prevention at ports: clearing the air. **Environmental Impact Assessment Review,** v. 24, n. 7-8, p. 749-774, 2004.

BAKONYI, S. M. C.; DANNI-OLIVEIRA, I. M.; MARTINS, L. C.; BRAGA, A. L. F. Poluição atmosférica e doenças respiratórias em crianças na cidade de Curitiba, PR. **Revista de Saúde Pública**, v. 38, n. 5, p. 695-700, 2004.

BRANIS, M.; SAFRANEK, J.; HYTYCHOVA, A. Exposure of children to airborne particulate matter of different size fractions during indoor physical education at school. **Building and Environment**, v. 44, n. 6, p. June 1246-1252, 2009.

CONCEIÇÃO, G. M. de S.; SALDIVA, P. H. N.; SINGER, J. da M. Modelos MLG e MAG para análise da associação entre poluição atmosférica e marcadores de morbi-mortalidade: uma introdução baseada em dados da cidade de São Paulo. **Revista Brasileira de Epidemiologia, v**. 4, n. 3, p. 206-219, 2001.

CORRAR, L. J.; PAULO, E.; DIAS FILHO, J. M. **Análise Multivariada**. 1. ed. – São Paulo: Atlas, 2009.

DOMINICI, F.; PENG, R. D.; BELL, M. L.; PHAM, L.; MCDERMOTT, A.; ZEGER, S. L.; SAMET, J. M. Fine Particulate Air Pollution and Hospital Admission for Cardiovascular and Respiratory Diseases. **JAMA**, v. 295, n. 10, p 1127-1134, 2006.

FUNDAÇÃO ESTADUAL DE PROTEÇÃO AMBIENTAL – FEPAM. **Relatório da Qualidade do ar 2001-2002**. Porto Alegre, 2002. 72 p.

GOUVEIA, N.; MENDONÇA, G. A. S.; LEON, A. P.; CORREIA, J. E. M.; JUNGER, W. L.; FREITAS, C. U.; DAUMAS, R. P.; MARTINS, L. C.; GIUSSEPE, L.; CONCEIÇÃO, G. M. S.; MANERICH, A.; CUNHA-CRUZ, J. Poluição do ar e efeitos na saúde nas populações de duas grandes metrópoles brasileiras. **Epidemiologia e Serviços de Saúde**, v. 12, n. 1, p. 29-40, 2003.

HAIR JR., J. F.; BLACK, W. C.; BABIN, B. J.; ANDERSON, R. E.; TATHAM, R. L. **Análise multivariada de dados**. Porto Alegre: Bookman, 2009.

MARTINS, L. C.; LATORRE, M. R. D. O.; CARDOSO, M. R. A.; GONÇALVES, F. L. T.; SALDIVA, P. H. N.; BRAGA, A. L. F. Poluição atmosférica e atendimento por pneumonia e gripe em São Paulo, Brasil. **Revista de Saúde Pública**, v. 36, n. 1, p. 88-94, 2002.

MCCULLAGH, P.; NELDER, J. A. **Generalized Linear Models**. Chapman and Hall, 1989.

SCHWARTZ, J. Air Pollution and Hospital

Admissions for Respiratory Disease. **Epidemiology**, *v*. 7, n. 1, p. 20-28, 1996.

SEGALA, C.; POIZEAU, D.; MESBAH, M.; WILLEMS, S.; MAIDENBERG, M. Winter air pollution and infant bronchiolitis in Paris. **Environmental Research**, v. 106, p. 96-100, 2008.

YORKS, J. E.; THOMPSON, A. M.; JOSEPH, E.; MILLER, S. K. The variability of free tropospheric ozone over Beltsville, Maryland (39N, 77W) in the summers 2004–2007. **Atmospheric Environment**, v. 43, p. 1827–1838, 2009.